Models and Measures of Beginning Teacher Quality

Linda P. Blanton
Florida International University
Paul T. Sindelar and **Vivian I. Correa**
University of Florida

Teacher quality is widely recognized as influencing student achievement and success in school. In this article, we consider various approaches to the assessment of teacher quality, including process–product observational measures, evaluation checklists, professional standards, large-scale surveys, and commercially available observation systems. We present examples of each from the special education literature, consider teacher education research genres for which each is appropriate, and evaluate each using a set of criteria that incorporates both practical and technical considerations. We advocate for multimethod approaches to teacher quality research and for more research relating what teachers know and do to what students learn, and we note that a stronger link between teachers and learners would allow for more rigorous evaluations of teacher preparation.

Drawing on analyses of a subset of teacher quality studies in which student outcomes were used as a dependent variable (Hess, 2001; Walsh, 2001), the No Child Left Behind Act (NCLB; 2001) challenged traditional concepts of good teaching by emphasizing content mastery and verbal ability and downplaying the importance of pedagogy. In turn, the belief that pedagogy is a less powerful determinant of student achievement than content mastery has led policymakers to propose alternatives to traditional teacher preparation. Thus, the NCLB encourages states to develop routes that move teachers into classrooms on "a fast-track basis" and includes in its definition of "highly qualified teachers" individuals enrolled in such alternative routes. The NCLB draws no distinction between secondary and elementary teachers, or between general and special education teachers, in spite of the fact that content mastery would seem to have relatively less relevance to the effectiveness of elementary and special education teachers and that pedagogy would have relatively more. Arguably, for most special educators, pedagogy would seem far more important as a determinant of achievement than mastery of the content they teach, which often involves basic skills.

Special education alternative routes are proliferating (Rosenberg & Sindelar, 2005), most probably in response to chronic shortages of special education teachers and the NCLB requirement that all teachers be highly qualified by the 2006–2007 school year. Furthermore, in a recent national survey of special education alternative programs, Rosenberg, Boyer, Sindelar, and Misra (in press) identified a small subset that has adopted NCLB-like, fast-track approaches as well. In programs of this sort, participants are provided limited training, in spite of their need for strategies for coping with significant learning and behavior problems. Widespread development of alternative routes and the existence of even these few fast-track alternatives led Rosenberg et al. to conjecture that special education has entered an era in which traditional standards for teacher preparation have given way to pragmatism. It is important to note that this transition has occurred in spite of limited empirical research on the efficacy of preparation alternatives, including traditional routes, and the equivocal findings that existing research has yielded (Nougaret, Scruggs, & Mastropieri, 2005; Sindelar, Daunic, & Rennells, 2004).

Thus, from teacher educators' perspective, research on the efficacy of alternative preparation routes seems more critical than ever before. For one thing, a policy direction is set, and its disregard for pedagogical training undermines what teacher educators believe about the scope and rigor of effective preparation. Policymakers also have raised the standard for credible evidence, so that changes to NCLB policy on teacher preparation and teacher quality will require not only more evidence but better evidence as well. Besides, to compete successfully for students in the entrepreneurial world in which teacher educators now work, guidance in designing effective alternative routes is essential. However urgent these considerations may be, rigorous and definitive research on the impact of teacher preparation cannot be easily—or cheaply—had. Perhaps the first and most important hurdle for teacher education researchers to surmount is to identify a credible and versatile measure of teacher quality, one that will garner the attention of both policymakers, who have set student outcomes as the gold standard for teacher quality, and teacher ed-

Address: Linda P. Blanton, EdD, College of Education, Florida International University, 11200 SW 8th St., ZEB 320, Miami, FL 33199-0001

ucators, who understand the difficulty of linking what they do first to the competence of their graduates and ultimately to the achievement of their graduates' students.

We recognize and appreciate the importance of linking what teachers do to what their students learn and how well they behave. Establishing links would allow teacher education researchers the opportunity to focus more specifically on linking the content of preparation to the competence of graduates. Measures of competence or quality (terms we use interchangeably in this paper) have not been used commonly in research on teacher preparation, in spite of the fact that concepts of teacher quality have evolved through an interesting series of representations.

Concepts of Teacher Quality

Regardless of one's purpose in doing so, defining teacher quality is no easy task. Reaching consensus on a definition, even among teacher educators and researchers, has proven elusive. As noted by Berliner (2005), "quality always requires value judgments about which disagreements abound" (p. 206). Definitions of high-quality teaching range in their focus from the actions of the teacher, to the knowledge a teacher possesses, to the creativity of the teacher. In recent years, however, leading researchers (Berliner, 2005; Fenstermacher & Richardson, 2005) have focused on the multidimensional nature of the concept and have defined teacher quality as encompassing two parts: (a) good teaching, meaning that the teacher meets the expectations for the role (e.g., holding degrees, using age-appropriate methods, upholding the standards of a field of study, and other attributes and practices), and (b) effective or successful teaching, meaning the results of the teacher's actions on student learning and achievement. In other words, one dimension in the absence of the other falls short of fully defining teacher quality.

The history of research on teaching and on the qualities that produce great teachers is relatively short. Although there were early studies in the 1940s, 1950s, and into the 1960s that focused on personal characteristics and experience variables, it was not until the late 1960s that researchers turned their efforts to exploring the link between specific teacher actions and student learning (Cochran-Smith & Lytle, 1990; Shulman, 1986). This process–product approach to research was based on behavioral psychology and child development, and although general education researchers initiated this line of research, some special education researchers conducted similar studies and contributed to the findings that influence teaching and teacher education today (Blanton et al., 2003). For example, effective teachers were found to (a) teach classroom rules and monitor expectations, (b) provide clear explanations and ample instructional time, (c) maximize the opportunity for students to respond during instruction and seatwork, (d) use a brisk pace to present lessons and present new material in small steps, and (e) provide regular feedback (Berliner, 1984; Christenson, Ysseldyke, & Thurlow, 1989; Englert, Tarrant, & Mariage, 1992; Good, 1979; Medley, 1978; Rosenshine, 1986; Shulman, 1986; Sindelar, Smith, Harriman, Hale, & Wilson, 1986).

In the 1970s, research began to address the complexities of teaching, classrooms, and schools, approaches to research that are referred to by different terms: learning-to-teach research, classroom ecology research, or interpretive research (Berliner, 1989; Doyle, 1983; Fenstermacher & Richardson, 2005; Kagan, 1992; Wideen, Mayer-Smith, & Moon, 1998). The literature grew rich with research on teacher planning/decision making (e.g., Reynolds, 1992), teacher thinking (e.g., Carter, 1990), teacher beliefs (e.g., Pajares, 1992), and novice versus expert teaching (Berliner, 1986), among other topics. As in their efforts in process–product research, special education researchers (e.g., Brantlinger, 1996; Fuchs, Fuchs, & Bishop, 1992; Nowacek & Blanton, 1996) borrowed from these new programs of inquiry and produced findings that added to the literature.

The knowledge base on teaching and understanding teacher quality continues to expand and change, focusing on both the good and the effective or successful dimensions of teacher quality. Currently, accountability and performance standards are dominating the teacher quality agenda, with accompanying changes in teacher education accreditation and teacher licensure, which are the major quality control mechanisms for the profession. The result of this focus is greater attention on such teacher attributes as experience, preparation and degrees, and certification (Rice, 2003).

Regardless of how difficult it is to encompass the concept of teacher quality, researchers need credible measures to build strong research programs. As we have argued, strong research programs are necessary both to guide teacher education program design and to inform policy. Although large-scale studies of teacher education program efficacy are under way (or complete) in general education (Darling-Hammond, 2000; Fenstermacher & Richardson, 2005; Humphrey & Weschler, 2005; National Commission on Excellence in Elementary Teacher Preparation for Reading Instruction, 2003), similar efforts are needed in special education. The purpose of this paper is to consider existing approaches to assessing beginning teacher quality and examine their utility for research in special education.

Evaluation of Five Approaches

We will consider five approaches to defining beginning teacher quality and measures associated with them: (a) process–product measures, (b) teacher evaluation checklists, (c) standards, (d) large-scale surveys, and (e) commercially available observations. We first discuss the general problem of teacher assessment and the particular problem created by the use of student achievement as a measure of teacher quality. Although some consider it a gold standard (Greenwood & Maheady,

1997; Walsh, 2001), we consider alternatives, or what Kennedy (1999) described as "approximations to indicators of student outcomes" (p. 345). For each model and measure, we consider the teacher education research genres (Kennedy, 1996) to which it applies and evaluate it against a set of criteria for technical adequacy and practicality.

The use of student outcomes, particularly achievement, as a measure of teacher quality enjoys strong support from both education professionals (Greenwood & Maheady, 1997) and the policy community (Walsh, 2001). The widespread and ready availability of standardized achievement test scores, the fruit of state policy on high-stakes assessment, has fostered interest in their use as an outcome measure in research on teacher quality. Although policymakers have always been interested in the impact of their initiatives on student learning, previously it was difficult to generate an adequately large database for analysis. High-stakes assessment has changed all that.

In a discussion of policy research measures, Kennedy (1999) commented on the difficulty of linking policy initiatives to student outcomes (especially when the outcome of interest is complex student learning) and described approximation methods. She argued that scores on standardized achievement tests, although first-order approximations of complex student learning, fail to represent it fully. According to Kennedy, classroom observations, another first-order approximation, may be better, particularly when the observations describe "the kind of intellectual work that teachers are asking of their students" (p. 346). However, observations suffer from other shortcomings. For one thing, there are no standard observation practices; for another, due to the time that observation requires, typically only small samples of teacher performance are obtained.

As a result, according to Kennedy, researchers tend to rely on second-level approximations, or "situated descriptions of teaching" (1999, p. 349). Second-level approximations include vignettes (with teachers' responses) and teachers' daily logs. Questionnaires and interviews constitute third-level approximations, and personal testimonies are fourth-level approximations. Although each level has advantages and disadvantages, generally, the more removed a measure is from complex student outcomes, the more likely it is that disadvantages will outweigh advantages. The advantages associated with third- and fourth-level approximations may be limited to the practical considerations of ease of administration and low cost. At those levels, technical adequacy may be compromised as well. In Kennedy's framework, most measures we consider in this article (process–product measures, checklists, and observations) are first-level approximations. The representations of teacher quality in the *Schools and Staffing Survey* (SASS; n.d.) and the *Study of Personnel Needs in Special Education* (SPeNSE; n.d.) are third-level approximations.

Kennedy's argument about complex student learning is only one criticism of the use of standardized test scores in assessing school or teacher quality. Of equal concern, especially for teachers, is the relationship between previous learning and test scores in any given year. Clearly, students who score poorly on standardized tests are likely to score poorly again in the future. Thus, teachers in classrooms with low-achieving students will compare unfavorably with colleagues who teach high-achieving students, regardless of the quality of their teaching. Teachers rightly complain that judgments based exclusively on scores from single administrations of achievement tests disadvantage teachers with large numbers of low-performing students.

In special education, the problem becomes more difficult, because classroom teachers and special educators share responsibility for educating most students with disabilities. Thus, determining which teacher is responsible for what learning may be impossible to do with any degree of precision or consistency. Furthermore, special education teachers' roles vary from school to school and, for some teachers, from student to student. A special education teacher may work with a single group for much of the day, work with several groups of students for short periods in a resource room, consult with some students' classroom teachers in planning accommodations and adaptations, or co-teach with a classroom teacher. With the possible exception of special education teachers in self-contained classes, the relationship between special education teacher quality and student outcomes is unclear and potentially tenuous.

As a result, in this paper, we consider models and measures of beginning teacher quality that are approximations of student outcomes. We wish a more definitive link were available between what special education teachers do and how much their students learn. At the same time, we recognize the importance of identifying approximations that are accurate and credible for teachers, researchers, and policymakers alike. Ultimately, of course, the two research traditions must merge so that teacher education practices may be linked to teacher quality and teacher quality to student outcomes.

Evaluation Criteria

The six criteria that we use to evaluate the models and measures of beginning teacher quality are *utility, credibility, comprehensiveness, generality, soundness*, and *practicality*. These criteria are represented as U, CR, CO, G, S, and P in Table 1. A plus (+) indicates that the criterion is regarded as a strength, a minus (–) indicates a weakness, and a plus–minus (±) indicates both strengths and weaknesses. The table lists specific examples of five general classes of models and measures, appropriate research genres, and the criteria used to evaluate each model.

Some criteria are pragmatic. For example, with regard to *utility,* we need to know whether models and measures have been used by other researchers. With a previously used measure, we can benefit from colleagues' experience, and their insight and advice may help us decide on appropriate measures for our own research. For *practicality,* also a pragmatic con-

TABLE 1. Models and Measures of Beginning Teacher Quality

Model	Examples	Research genre					Criteria used to evaluate models					
		1	2	3	4	5	U	CR	CO	G	S	P
Process–Product	COKER (Stallings, 1980)		√		√	√	+	±	-	+	+	-
Teacher evaluation checklists	Englert, Tarrant, & Mariage checklists (1992)[a]		√		√	√	-	±	+	+	-	-
	Stanovich & Jordan (1998)[b] and Haager et al. (2003)[b]		√		√	√	±	±	+	±	+	-
Standards	CEC Knowledge and Skills[c]		√	√		√	±	+	+	+	-	-
	INTASC[c]											
Representations of teacher quality in large-scale surveys	SASS	√		√			+	-	-	+	+	+
	SPeNSE											
Commercially available observation	PRAXIS III		√		√	√	±	+	+	+	+	-

Note. 1 = searches for factors that influence student outcomes; 2 = comparative studies of licensed and unlicensed teachers; 3 = follow-up surveys; 4 = experiments; 5 = case studies of change over time; U = utility; CR = credibility; CO = comprehensiveness; G = generality; S = soundness; P = practicality; COKER = *Classroom Observation Keyed for Effectiveness Research;* CEC = Council for Exceptional Children; INTASC = Interstate New Teacher Assessment and Support Consortium; SASS = *Schools and Staffing Survey* (SASS, n.d.); SPeNSE = *Study of Personnel Needs in Special Education* (SPeNSE, n.d.).
[a]Designed for student evaluation. [b]Designed for research. [c]Neither CEC nor INTASC currently offers an assessment process.

sideration, we need to understand costs, training requirements, and the developmental work required to adapt an existing model or measure for our own purposes. All other considerations being equal, *inexpensive*, *easy-to-master*, and *readily adaptable* are preferred qualities.

Some of our criteria are technical in nature. We use s*oundness* to refer to the extent to which a measure is reliable and valid, and *credibility* to refer to face validity, which we separate from soundness to highlight its relativity and subjectivity. Although models and measures must be credible to the researchers using them, we are equally concerned with the credibility of a given model or measure for other stakeholder groups—most importantly, teachers, administrators, policymakers, and families. In this sense, credibility may be inferred from what we know about how a model or measure was developed and validated. We may infer credibility for stakeholder groups on the basis of the extent to which they were involved in the development or validation process.

Generality and *comprehensiveness* refer to a model's theoretical foundation. Generality requires us to consider how well a single model of beginning teacher quality represents the full range of contexts in which a special education teacher may work. Does the model fairly represent the work of co-teachers, consulting teachers, resource room teachers, and teachers in self-contained classes? Does the model fairly represent the work of teachers of students with high-incidence disabilities as well as students whose disabilities are more significant? Models that allow for comparability across contexts simplify the aggregation of research findings. Comprehensiveness is derived from the richness and breadth of the model or measure. A better model or measure taps knowledge and dispositions in addition to skills. A better model or measure includes management skills, reflection, and decision making in addition to discrete teaching performances. Finally, a better model or measure incorporates the work that teachers do with each other, families, and communities.

Research Genres

Kennedy (1996) described five traditions in teacher education research and considered for each genre the teacher education elements studied, the measures typically used, and the logic underlying each. The five genres are (a) identification of factors that influence student outcomes, (b) comparative studies of licensed and unlicensed teachers, (c) follow-up surveys, (d) experiments, and (e) case studies of change over time. The genres are described in greater detail in the paragraphs that follow.

Factors That Influence Student Learning. Studies of factors that influence student learning commonly use large-scale multiple regression models to analyze the statistical relationships between a set of predictor variables (including

teacher qualifications and teacher education variables, e.g., licensure) and a criterion variable (e.g., reading achievement). Such studies focus on the "policy parameters" (Kennedy, 1996, p. 124) of teacher education (e.g., number of required credits), and student achievement is typically the criterion of interest. An effort is made to identify variables that contribute to achievement and those that do not. In spite of limitations with the genre, studies of factors that influence student learning have the distinct advantage in the current policy context of using an achievement criterion.

In special education, researchers have few large-scale, nationally representative databases available to them that permit analysis of teacher quality. In education databases (generally), the sample of special education students or teachers, if identifiable, is not sufficiently large to permit refined statistical analyses. For example, in the 2003 administration of the National Assessment of Educational Progress (Perie, Grigg, & Donahue, 2005), nearly 190,000 students were assessed in reading, but fewer than 10,000 of them were identifiable as students with disabilities (by virtue of requiring testing accommodations). Similarly, the limited number of special education teachers in the SASS (n.d.) and Teacher Follow-up Survey (TFS; n.d.) samples has prevented Boe and his colleagues from computing for special education teachers the same array of estimates they compute for the general education sample. For example, in a recent study of teacher attrition and transfer, Boe, Cook, and Sunderland (2006) aggregated three SASS and TFS administrations but nonetheless were unable to estimate reliably the number of special education teachers who moved from one school to another and the number who left due to dissatisfaction with teaching.

Comparisons. Comparisons of licensed and unlicensed teachers typically involve observations of classroom practice or performance on teacher assessments. Because differences favoring fully qualified teachers are expected, studies of this genre test the value of teacher preparation explicitly. However, one problem with the logic underlying comparative studies is that teacher education is treated as a consistent and uniform phenomenon, which it is not. Furthermore, comparative studies also presume substantial differences in preparation, although most unlicensed teachers typically have completed at least some teacher preparation.

Follow-up Studies. Researchers operating within this genre presume that teachers themselves are reliable sources of information about their knowledge and skills and how these were acquired. Such studies may focus on components of teacher education and thereby allow for more precision than either of the first two genres, in which teacher education is considered to be a uniform and consistent intervention. Follow-ups that involve telephone, paper-and-pencil, or Web-based surveys can be administered widely for little cost. With large samples that permit stratification, teacher groups can be differentiated on key variables (e.g., graduates of 4-year vs. 5-year programs). Follow-up studies typically are conducted with graduates of a single teacher education program and are most useful for faculty there.

Experiments. In experimental studies of teacher education, a skill is taught in different ways with different groups, and differences in skill performance are attributed to differences in teacher education pedagogy. Experimental studies enjoy several advantages, including clear focus on teacher education components and assessment of outcomes (e.g., the skill being taught). However, such studies focus on training discrete, narrowly defined skills, which are part of—but not the sum of—teacher quality. Absent from experimental research are cognition, reflection, and decision making, the elements thought to make effective teaching a coherent whole.

Because special education has philosophical roots in positivism and special education researchers have methodological skill in designing studies with small samples, experiments are more common in teacher education research than in other genres (Tulbert, Sindelar, Correa, & La Porte, 1996). However, the problem of tight focus on observable actions is evident in special education teacher education research, perhaps because of our deep roots in behavioral psychology. When experiments have been conducted, single-subject designs have been used. Although such designs allow researchers to demonstrate control over dependent measures, these dependent measures are limited to observations of discrete actions.

Case Studies. In case studies, teacher candidates are examined at the beginning and end of their programs, and possibly more often. Differences on these assessments are used to describe the process through which a teacher develops. Candidates' knowledge, attitudes, and beliefs may be assessed. If cost were no consideration, observations of classroom practice also could be used within this genre. In good case study research, theory is used to generate and organize questions and to suggest directions for change.

Case study methods, more so than follow-up studies, seem well suited to the special education context. Good follow-up studies (like the SPeNSE) require large samples, which are hard to constitute in special education. For one thing, state-to-state variation in certification structure limits our ability to lump teachers together into a nationally representative sample. Second, some specializations within special education (e.g., deaf and hard of hearing) require mastery of at least some content that is unique to the specialization.

Models and Measures

The five traditions of assessing beginning teacher quality are (a) empirical representations of effective practice derived from process–product research; (b) more complete and holistic representations, exemplified by checklists developed by

Englert and her colleagues (1992) and others (Stanovich & Jordan, 1998); (c) standards; (d) representations of effective practice from large-scale surveys (e.g., SASS, SPeNSE); and (e) observation systems for classroom teachers such as the PRAXIS III, developed and published by the Educational Testing Service (ETS). We next weigh these traditions against our six evaluation criteria and consider the genres for which each would be appropriate. In doing so, we cite research in which specific examples of the practices were used. We urge readers to consider these studies illustrative and not exhaustive and to bear in mind that the measures we do consider are better thought of as prototypes than as exemplars.

Process–Product Observational Measures

In the typical process–product study, teachers are observed at work in their classrooms. Teaching and classroom interactions typically are described in a series of low-inference behavioral categories, often mutually exclusive and exhaustive, so that any event may be coded in only one way. The manner in which the stream of classroom events is parsed reflects an empirical or theoretical conception of teaching. Code frequencies or durations are aggregated across teachers and related to achievement measures. Relationships between patterns of classroom performance (or interactions) and student outcomes are determined statistically.

To illustrate, in Algozzine, Morsink, and Algozzine's (1988) study of instruction in self-contained special education classrooms, the researchers used the *Classroom Observation Keyed for Effectiveness Research* (COKER). Medley, Coker, and Soar (1984) described the COKER as an objective, low-inference process for observing the ongoing flow of student–teacher interaction. Based on its history of use in general education process–product research and information from the manual, Algozzine et al. judged the COKER to be technically adequate for their purposes. The system requires trained observers who code all of the *keys* they observe in a given time period. In the COKER lexicon, keys are statements describing discrete teacher actions—what others might call competencies, performances, or behaviors. For example, one key under Learner Reinforcement and Involvement is "maintains environment in which students are actively involved, working on task" (Medley et al., 1984, p. 162). The COKER is a complex system, and Algozzine et al. used three of its Competency Dimensions: (a) Instructional Strategies, Techniques, or Methods (7 keys), (b) Communication With Learners (5 keys), and (c) Learner Reinforcement and Involvement (5 keys). On the basis of these COKER observations, Algozzine et al. reported that the teachers in their study performed adequately, but not differently, regardless of the classifications of their students.

Process–product measures like the COKER seem well suited to comparison studies of licensed and unlicensed teachers and longitudinal studies of change, although using such complex systems would be costly and labor-intensive. These measures can also be used in experiments, as in Stallings's (1980) work on beginning reading instruction. In this study, teachers were assessed before and after an intervention designed specifically to affect how they allocated time across activities.

We have alluded to the high cost of repeated administrations of process–product measures and other factors that limit their practicality. For example, extensive training is required for COKER observers, and the need to repeat training over the duration of a longitudinal study further diminishes its practicality. Furthermore, it is uncertain whether an established system will be sensitive to the changes that a particular program is intended to produce. The credibility of a process–product measure like the COKER may derive from professional consensus, research on effective teaching, or theory. COKER keys, originally developed through professional consensus, were validated in subsequent research (Medley et al., 1984). Its use in special education classrooms required a leap of faith by Algozzine et al. (1988), but the generality of the system was borne out by its utility to the authors. As a limitation, conceptions of teacher quality derived from the COKER—or from process–product measures in general—are based on observations of teachers' actions and fail to tap other dimensions of what we know to be complex performance.

Overall, process–product measures have strengths and weaknesses. Foremost among their strengths is the potential for highly reliable measurement of the relationships between items on the observation system and key criterion variables, such as achievement. Among the weaknesses of process–product measures is the reliance on teachers' actions to the exclusion of internal events available through interviews, logs, and other measures. Their use also may be impractical, particularly when extensive training is required for reliable administration, or when research designs necessitate repeated observations over time.

Teacher Evaluation Checklists

In 1992, Englert, Tarrant, and Mariage described a series of detailed, moderate-inference checklists that they had developed for evaluating field experience students. Taken together, Englert et al.'s checklists constitute a rich, detailed model of beginning teacher quality. In this section, we consider both the original checklists and a research adaptation developed by Stanovich and Jordan (1998) for their study of the relationship of teachers' and principals' beliefs about inclusion and effective teaching practices.

The first four checklists described by Englert et al. were derived from process–product relationships and covered classroom management (15 items), time management (10 items), lesson presentation (27 items), and seatwork management (9 items). All items were scored on a 1-to-5 scale, ranging from *needs work* (1) through *satisfactory* (2–3) to *excellent* (4–5). The authors offered no guidance about how long an observation must be conducted before reliable judgments can be made, nor did they provide other evidence of technical adequacy.

To these process–product checklists, Englert et al. added additional items that "involve analyses of the qualitative dimensions of instruction and the social contexts in which students are instructed" (1992, pp. 69–70). This enriched process–product conceptions by adding items derived from the following four principles of effective teaching: (a) instruction should be embedded in meaningful and purposive contexts; (b) classroom dialogue may be used to promote self-regulated learning; (c) teachers must demonstrate responsiveness to students' instructional needs and interests; and (d) in classroom learning communities, "student-to-student and teacher-to-teacher discourse . . . foster deeper conceptual understandings" (p. 80). To incorporate these constructivist principles, Englert et al. added the *Observation Checklist for Examining the Contexts for Higher-Order Learning,* which corresponds to the four principles: meaningful contexts (4 items), classroom dialogues (11 items), responsive instruction (8 items), and classroom community (5 items).

Stanovich and Jordan (1998) adapted Englert et al.'s checklists by identifying items that would be most readily and predictably observed in a half day, using 8 of Englert et al.'s items to constitute a classroom management scale, 8 to create a time management scale, and 11 to create a lesson presentation scale. They also added 4 items to assess the degree of inclusion. Trained observers rated teachers' performance on these 31 items after a half-day observation. Items were scored as *consistent, inconsistent,* or *not observed.* Total scores were used as the criterion measure of effective teaching. All teachers were rated by two observers, and agreement between observer pairs averaged nearly 80%. Also, *The English-Language Learner Classroom Observation Instrument* (Haager, Gersten, Baker, & Graves, 2003) was designed specifically for observing literacy instruction by teachers of English-language learners. This instrument, whose roots also may be found in process–product research, was validated for research purposes and has value for classrooms with culturally diverse and English-language learners.

Checklists of this sort lend themselves to the same kinds of teacher education studies as do process–product observational measures, to which they are closely akin. These seem appropriate for use in comparative studies, experiments, and case studies of change. Stanovich and Jordan's adaptation and use enhance the utility of the original checklists, which seem impractically long and elaborate for research purposes. Furthermore, Stanovich and Jordan demonstrated that their abbreviated versions can be used reliably and that short-form total scores were related to two criterion measures: teacher attitudes and school culture.

The Englert et al. paper was the most frequently cited paper to appear in *Teacher Education and Special Education* through 1995 (Tulbert et al., 1996), which suggests its strong credibility to an audience of teacher educators. By intention, the full-length checklists are more comprehensive than process–product measures, as they include considerations of contextual factors, interactions, and community, which are notably missing from behavioral observation systems and from the abbreviated version used by Stanovich and Jordan. The checklists have wide applicability in assessing teachers of students with high-incidence disabilities. Englert's important work advanced special education thinking about what constitutes effective teaching. The ideas she and her colleagues introduced a decade ago seemed quite radical indeed. At a practical level, however, the checklists have never been widely used for research purposes, Stanovich and Jordan's work notwithstanding.

Standards

The Council for Exceptional Children (CEC) began promulgating teaching standards in the early 1990s and in 2001 published a revised edition of *The CEC Standards for the Preparation of Special Educators*. This document begins with narrative descriptions of 10 content standards: foundations, development and characteristics of learners, individual learning differences, instructional strategies, learning environments and social interactions, communication, instructional planning, assessment, professional and ethical practice, and collaboration. Each content standard is then described in terms of the knowledge and skill competencies it comprises.

Fifty-four knowledge and 72 skill statements make up the common core. Additional sets describe generic practice with students with high-incidence (individualized general curriculum) and severe (individualized independence curriculum) disabilities. Specialized practice is represented in six categorical areas and two other areas (early childhood and transition specialist) defined by the age levels of the students served and the nature of appropriate programming. A trainee who completes generic preparation for teaching students with high-incidence disabilities is expected to demonstrate proficiency on 126 competencies in the core as well as 42 knowledge and 47 skill statements in the individualized general-curriculum-referenced standards. A teacher preparing to work with students with specific learning disabilities (SLD) must demonstrate proficiency on 174 competencies: the core plus 27 knowledge and 21 skill statements specific to SLD.

The CEC's knowledge competencies are written as general, descriptive phrases. One statement from the common core—development and characteristics of learners—reads, "Educational implications of characteristics of various exceptionalities" (CEC, 2001, "Common Core," p. 1). Another from the specialized knowledge base in mental retardation—development and characteristics of learners—is "causes and theories of intellectual disabilities and implications for prevention" (CEC, 2001, "Mental Retardation/Developmental Disabilities," p. 2). Skill standards start with verbs and are like knowledge statements in their generality. In fact, some skill descriptions are so general as to belie their use as categorical standards. For example, teachers of students with SLD are expected to "use specialized methods for teaching basic skills" (CEC, 2001, "Learning Disabilities," p. 1). Others are more

precisely described and more clearly associated with a particular categorical area; for example, "demonstrate appropriate body mechanics to assure student and teacher safety in transfer, lifting, positioning, and eating" (CEC, 2001, "Independence Curriculum," p. 3) seems quite specific to individualized independent-curriculum-referenced standards, learning environments, and social interactions.

Whereas CEC knowledge and skill items are precisely defined, standards developed by the Interstate New Teacher Assessment and Support Consortium (INTASC) are fewer in number and more broadly conceived. INTASC standards are organized by the principle to which they are related. By virtue of the CEC's effort to align their standards with those of INTASC, the 10 INTASC principles are roughly analogous to the CEC's 10 content standards.

The title of INTASC's document for teachers working with students with disabilities, *Model Standards for Licensing General and Special Education Teachers of Students With Disabilities: A Resource for State Dialogue* (2001), hints at its organization. Every principle is elaborated into standards for general and special education teachers and additional standards for special education teachers only. INTASC standards, first released in 1992, were designed for compatibility with standards for accomplished practice promulgated by the National Board of Professional Teaching Standards (NBPTS). The special education initiative began in 1997. These standards, which were developed by a committee of general and special education teachers and teacher educators, include knowledge, skills, and dispositions that build on and are organized by the core principles. The purpose of these standards is to differentiate general from special education teachers' roles, with reference to (a) content (Principle 1); (b) pedagogy (Principles 4–10); (c) knowledge of students with disabilities (Principles 2 and 3); and (d) contexts (Principle 10). There are 49 standards for both general education and special education teachers and an additional 49 for special education teachers.

INTASC (2001) standards (a) emphasize that "teaching and learning are dynamic and interactive processes" (p. 2); (b) are responsive to students' contexts; and (c) encourage users to take standards as a whole "to convey a complete picture of the acts of teaching and learning" (p. 2). Unlike the CEC standards, INTASC knowledge, skills, and dispositions are not differentiated by teachers' roles or students' disability classifications. The statements are written in paragraph-length narratives of complete sentences. Typically, a principle is broken down into elements, which are elaborated in the standards. For example, for Principle 3, "the teacher understands how students differ in their approaches to learning and creates instructional opportunities that are adapted to diverse learners" (p. 17), the general and special education teacher standards include (a) building awareness of disability and respect for students with disabilities, (b) recognizing that students with disabilities make up a heterogeneous group, (c) understanding families' perspectives on disabilities, and (d) recognizing that some differences may be mistaken for disability.

Because neither the CEC nor INTASC standards offer an assessment process (although assessments are in the works), it is impossible to speak to the issue of their technical soundness or their utility as an approximation for student achievement. At present, standards seem most useful for guiding the development of surveys of graduates in follow-up studies and in interviews used in longitudinal studies of change, perhaps associated with accreditation reviews. Standards have rarely been used as outcome measures in teacher education research; their use in Nevin, Thousand, Parsons, and Lilly (2000) is the only instance we found in the special education literature. However, both the CEC and INTASC standards have the decided advantage of being fully comprehensive and general by design—the CEC standards in a formal sense by differentiating knowledge and skill items and by roles. Both conceptions include important work that teachers do outside the classroom. In the CEC's collaboration standard and INTASC's Principle 10, "The teacher fosters relationships with school colleagues, families, and agencies in the larger community to support students' learning and well being" (INTASC, 2001, p. 37). Both the CEC and INTASC standards were developed over iterations with input from key stakeholders.

For the moment, standards have limited potential as outcome measures in special education teacher education research, except as a guide to survey or interview development in follow-up or longitudinal research. At the same time, the conceptions of beginning teacher quality represented in these standards are detailed, coherent, and complete. The standards do represent contemporary professional thought but, unlike process–product measures, lack empirical connection to student outcomes.

Large-Scale Surveys

Questions in the SASS (n.d.) and SPeNSE (n.d.) teacher surveys also constitute representations of beginning teacher quality. The SASS has been administered five times since 1987 by the National Center on Educational Statistics (NCES), and Boe and his colleagues have used the SASS data in analyses of special education teacher supply and demand (Boe, Bobbitt, & Cook, 1997; Boe, Cook, Bobbitt, & Terhanian, 1998; Boe, Cook, Kaufman, & Danielson, 1996). The SASS sample taps the universe of public and private schools in the United States. The SPeNSE survey was administered once to a nationally representative sample of general and special education teachers.

SASS. The SASS *Teacher Questionnaire* asks teachers to specify demographics, educational backgrounds, certification(s), and years of experience. Additional questions tap (a) length of practice teaching, (b) first-year duties and supports, (c) mentoring, and (d) professional development. The questionnaire assesses job satisfaction, attitudes and perceptions about support, influence in school, school safety, and behavior. Teachers also are asked how well prepared they felt in their first year of teaching for management, instructional

methods, technology, lesson planning, assessment, and selection and adaptation of instructional materials. Other questions focus on professional development.

SPeNSE. In the SPeNSE teacher survey, teachers were asked about preservice preparation and to indicate the number of hours of professional development they received over the previous 12 months in each of 27 areas. Teachers also were asked to indicate their degree of agreement on a Likert-type scale with statements such as "I am skillful in planning effective lessons," and "I am skillful in teaching reading or prereading skills." Thus, for the 27 preparation areas, teachers indicated any preservice training, specified hours of professional development, and judged their degree of mastery.

SPeNSE included a second set of questions about professional development. Teachers were asked whether they had a personal professional development plan and whether they had participated in any of 12 professional development activities over the past year. They indicated their hours spent in professional development and benefits of these experiences (e.g., Improved your effectiveness as a teacher? Been responsive to your professional development needs?). This section of the survey ended with six more questions related to mentoring, contacts with teachers and other education professionals, reading professional journals, and association membership.

Representations of beginning teacher quality, like SASS and SPeNSE, clearly are intended for use in follow-up survey research. Their utility is evident, and these surveys have been used with both general and special education teachers and across special education contexts. The entire research genre is practical in that extensive data may be generated relatively inexpensively and relatively quickly. Because SASS and SPeNSE have been validated by use in previous research, these surveys are presumed to be technically sound, although the conceptions of beginning teacher quality that can be inferred are sketchy and incomplete relative to other potential measures (e.g., standards). Generally, the credibility of surveys like these is limited by the self-report format and its potential for inaccuracy and bias.

Commercially Available Observation Systems: PRAXIS III

PRAXIS III (Dwyer, 1993, 1994) is "a system for assessing the teaching skills of beginning teachers" (Dwyer, 1998, p. 163). (PRAXIS I is a test of enabling skills such as reading, writing, and arithmetic, and PRAXIS II is a test of subject matter knowledge and teaching principles.) The 19 PRAXIS criteria, which are organized into four domains, were developed from research (Reynolds, 1992), job analyses, and a multistate validity study. The criteria were piloted in the field and refined in collaboration with practicing teachers. During its development, PRAXIS III went through five iterations.

The development process, begun in 1987 and completed in 1993, involved (a) establishing an underlying conception of teaching, (b) developing a plan for defining teaching, and (c) linking this definition to assessments. The underlying conception of teaching emphasizes the importance of action and decision making and the consideration of individual, school, and community contexts. Because learning is presumed to involve the active construction of knowledge, assessments must take place in classrooms. Teachers are afforded opportunities to explain their actions, and scoring allows for the reality that good teaching can take many forms. Skilled professionals are thought to make the best assessors. The 19 criteria are organized into four domains: (a) organizing content knowledge for student learning, (b) creating an environment for student learning, (c) teaching for student learning, and (d) teacher professionalism.

Dwyer (1998) asserted that these criteria establish "a vision of teaching . . . derived from working closely with teachers themselves . . . relevant to teachers' own practice and concerns . . . [and] informed by the theoretical and policy perspectives of other educators and researchers" (p. 172). PRAXIS III involves three data collection processes: (a) direct observation of classroom practice (in which a running narrative is kept), (b) written materials (class and teacher profiles and a lesson plan), and (c) interviews (before and after the observation) related to the lesson.

Trained assessors observe teachers as they teach a lesson of their choice to a group of their choice. Using the full record of evidence—the two profiles, lesson plan, running observational record, and interview protocols—assessors rate teachers on the 19 criteria. The scale used is from 1.0 to 3.5; a rating of 2.0 represents minimally satisfactory performance. Scoring is guided by a rubric linked to the nature of the evidence. Assessor training, which requires 5 days, is considered essential for identifying evidence relating to criteria and to using evidence to reach judgment.

Observation systems like PRAXIS III may be used in comparative studies and in longitudinal studies of change. The generality of its criteria preclude its use in experiments unless ratings are conducted on a pre–post basis (as for process–product measures) and the intervention is designed specifically to affect performance on one or more criteria. The conception of teacher quality is highly credible, given the systematic manner with which it was developed and the participation of key stakeholders throughout. However, as an assessment for classroom teachers, PRAXIS III made no special adaptations for special education practice.

PRAXIS III has been used in one special education study (Sindelar et al., 2004). In this study, the system worked well with a sample of special education teachers, prompting the authors to conclude that PRAXIS III "provide[d] a clear and coherent picture of the competence of teachers . . . despite the fact that Praxis III was designed to assess general education teachers" (p. 222). They noted that the pattern of performance across the 19 criteria supported this assertion in the sense that their sample of special education teachers performed relatively well where expected and relatively less well

where expected (e.g., "extends students' thinking"). Ratings on six criteria and two domain summary scores differentiated graduates of three distinct teacher education program types.

Thus, PRAXIS III rates high marks on utility and credibility. Furthermore, the richness of the record of evidence creates a comprehensive picture of beginning teacher quality. With regard to soundness, Dwyer (1998) emphasized its construct validity and argued that construct validity was the most important consideration for teacher observation systems. Furthermore, ETS developed PRAXIS to market to states as a legally defensible process for licensing beginning teachers. The reliability of assessors' ratings is implied by the extensiveness of their training. However, PRAXIS III administration is highly costly and labor-intensive. Training is also costly and, for longitudinal studies, would need to be repeated for new assessors. Although the picture of teaching competence derived from PRAXIS III observations is rich and sound, the system may be impractical for some purposes.

Summary and Recommendations

Teacher quality means different things to different people. Moreover, different people use models and measures of teacher quality differently depending on their purposes. On one hand, a researcher may be willing to use measures that take time and are more difficult to administer because of an interest in understanding deeply many dimensions of teacher quality. On the other hand, a policymaker may want to acquire information quickly and efficiently and thus will call on measures that will accomplish this purpose. For example, defining beginning teacher quality as being fully certified is likely to have greater credibility among policymakers than among researchers. Even within the community of researchers who study teacher quality, there is no single definition or measure for beginning or experienced teachers, either in general education or in special education.

As inquiry into teaching and teacher education has grown and matured, in both general education and special education, models and measures of teacher quality have evolved. Yet, in special education, research on teaching (e.g., process–product studies) has focused most often on teachers of students with high-incidence disabilities, and, in this article, we focused on the same group. Our analysis persuades us that the models and measures we have presented are useful in studying this subset of special education teachers. A review of the literature related to beginning teachers serving culturally diverse and English-language learners (Blanton et al., 2003) led us to conclude that the models and measures are also useful for examining teacher quality in this area.

In this paper, we identified classes of models and measures, presented illustrative examples of each class, considered research genres for which each class would be appropriate, and discussed their merits using evaluation criteria. These analyses, as summarized in Table 1, lead to a single, irrefutable conclusion: The superiority of one model over another depends on the purpose and context of its use. For most purposes, the best approach would be to pick and choose from several models—and, indeed, using multiple measures in teacher education research is encouraged by most authorities (Wilson, Floden, & Ferrini-Mundy, 2002). In special education, there is a great need to accelerate research on beginning teacher quality by drawing on models and measures set forth in this paper. Our analysis suggests that special education teacher education research should be guided by these considerations:

1. Use multiple research traditions in conducting research on teacher quality and multiple measures associated with those traditions. With the exception of process–product research, special education has produced only a handful of research studies drawn from other research traditions. This fact alone calls out to special educators to expand research on teacher quality to include programs of research focused on understanding the complexity of teachers' actions and interactions with students and contexts.
2. Get the attention of policymakers by producing compelling research findings and by linking measures of teacher quality to student outcomes. Because policymakers need measures of teacher quality to communicate with the public, we offer two recommendations to special education researchers. First, it is critical that the special education research community take research on teacher quality seriously. We must accumulate findings that policymakers find credible and that distinguish our field from general education.

   A second facet of this problem derives from the fact that the difficulties special education researchers have in linking teachers' competence to student achievement may be too esoteric for policymakers to appreciate fully. For example, because the vast majority of students with disabilities spend the majority of the school day in general education classes, more than one teacher contributes to their academic growth (unlike, say, a typical third grader). In the typical special education context, determining which teacher is responsible for how much growth may be a practical impossibility. Furthermore, even if researchers could successfully parse achievement by teacher, their ability to relate achievement growth to teacher quality would be diluted by the smaller effect sizes associated with individual teachers. Second, special education teachers work with students whose abilities and rates of progress vary widely, and they are un-

likely to have an adequate number of students at roughly the same level of ability so as to make prediction feasible—as would a typical third-grade teacher, for instance. Furthermore, whereas in a typical elementary school there are three or four third-grade teachers, each working with children at about the same ability level, schools are likely to have one or two special education teachers, who probably divide their work by grade level or by the severity of their students' disabilities. To find four special education teachers all working with children at the same level of ability, then, researchers would have to be in four schools, and even then there would be no guarantee that children in those four schools were served in the same delivery model, or for the same length of time, or in the same curriculum. Thus, the difficult logistics and high cost of conducting research in schools are multiplied for special education researchers. Our job is to educate policymakers about the complexity of teaching and learning in special education contexts so that policy becomes something more than a standard solution imposed on distinct problems.

3. Validate assessments based on teaching standards. We noted earlier in our paper that assessments have not been developed, or are in very early stages of development, for the CEC and INTASC standards. Other groups, most notably the NBPTS, are farther along on the validation process, and we believe the CEC and INTASC would be well advised to follow NBPTS's lead by developing research programs to validate their assessment processes. We found three validation studies of the NBPTS assessment process (Bond, Smith, Baker, & Hattie, 2000; Goldhaber & Anthony, 2004; Vandevoort, Amrein-Beardsley, & Berliner, 2004). In these studies, NBPTS-certified teachers were compared with non-NBPTS-certified teachers (Goldhaber & Anthony, 2004; Vandevoort et al., 2004) or to teachers who had applied for certification but were turned down (Bond et al., 2000). Bond and colleagues studied teachers who applied for certification in either the early adolescence language arts or middle childhood generalist category, whereas both Goldhaber and Anthony (2004) and Vandevoort et al. (2004) studied elementary teachers. These researchers found that students of NBPTS-certified teachers typically outperformed students of comparison group teachers on measures of academic achievement (Goldhaber & Anthony, 2004; Vandevoort et al., 2004) or the quality of student work (Bond et al., 2000). The teachers in the Bond et al. study consistently outperformed comparison teachers on 13 measures of teaching excellence.

The questions such studies address are important ones, but knowing that the NBPTS (or INTASC or CEC) assessment process validly differentiates good teachers from outstanding ones has limited utility in the context of teacher shortages. In special education, it may be more important for assessment systems to reliably differentiate basically competent from incompetent teachers. This distinction seems appropriate for INTASC standards, which are designed for new teachers, but it is unclear as to where the CEC stands on this issue. In our judgment, because of the leadership role that the CEC plays in the field of special education and until shortages are addressed, that organization should concentrate on developing an assessment process for novice teachers and establishing its validity in identifying basically competent ones.

Research on teacher quality is as challenging as it is important. Policy is in place that most special education teacher educators regard as detrimental to teachers' professional development and students' success in school. A high standard has been set for the credibility of evidence that policymakers will consider. Historically, teacher education researchers have endeavored to link variation in preparation to variation in teacher competence and, in doing so, did not restrict outcomes to a student achievement standard. In fact, Kennedy (1999), among others, has argued that direct observation is a better measure of teacher quality than achievement test scores, provided that the outcome of interest is complex student learning. Nonetheless, in the current policy context, scholars would be naive to ignore student outcomes. Thus, future research in our field must focus on the validation of measures of teacher quality. Only then will researchers have the tools they need to link preparation variables to credible measures of teacher quality. Only then will they garner the attention of policymakers.

REFERENCES

Algozzine, B., Morsink, C. V., & Algozzine, K. M. (1988). What's happening in self-contained special education classrooms? *Exceptional Children, 55,* 259–265.

Berliner, D. C. (1984). The half-full glass: A review of research on teaching. In P. L. Hasford (Ed.), *Using what we know about teaching* (pp. 51–77). Alexandria, VA: Association for Supervision and Curriculum Development.

Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher, 15*(7), 5–13.

Berliner, D. C. (1989). The place of process–product research in developing the agenda for research on teacher thinking. *Educational Psychologist, 24,* 325–344.

Berliner, D. C. (2005). The near impossibility of testing for teacher quality. *Journal of Teacher Education, 56,* 205–213.

Blanton, L., Sindelar, P. T., Correa, V., Hardman, M., McDonnell, J., & Kuhel,

K. (2003). *Conceptions of beginning teacher quality: Models for conducting research* (COPSSE Document No. RS-6). Gainesville, FL: University of Florida, Center on Personnel Studies in Special Education.

Boe, E. E., Bobbitt, S. A., & Cook, L. H. (1997). Whither didst thou go? Retention, reassignment, migration, and attrition of special and general education teachers from a national perspective. *The Journal of Special Education, 30,* 371–389.

Boe, E. E., Cook, L. H., Bobbitt, S. A., & Terhanian, G. (1998). The shortage of fully certified teachers in special and general education. *Teacher Education and Special Education, 21,* 1–21.

Boe, E. E., Cook, L. H., Kaufman, M. J., & Danielson, L. (1996). Special and general education teachers in public schools: Sources of supply in national perspective. *Teacher Education and Special Education, 19,* 1–16.

Boe, E. E., Cook, L. H., & Sunderland, R. J. (2006). *Teacher turnover in special and general education: Exit attrition, teaching area transfer, and school migration* (Research Report No. 2006-TSDQ3). Philadelphia: University of Pennsylvania, Center for Research and Evaluation in Social Policy.

Bond, L., Smith, T., Baker, W. K., & Hattie, J. A. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study.* Greensboro, NC: The University of North Carolina at Greensboro.

Brantlinger, E. (1996). Influence of preservice teachers' beliefs about pupil achievement on attitudes toward inclusion. *Teacher Education and Special Education, 19,* 17–33.

Carter, K. (1990). Teacher's knowledge and learning to teach. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 291–310). New York: Macmillan.

Christenson, S. L., Ysseldyke, J. E., & Thurlow, M. L. (1989). Critical instructional factors for students with mild handicaps: An integrative review. *Remedial and Special Education, 10*(5), 21–31.

Cochran-Smith, M., & Lytle, S. L. (1990). Research on teaching and teacher research: The issues that divide. *Educational Researcher, 19*(2), 2–11.

Council for Exceptional Children. (2001, August). *The CEC standards for the preparation of special educators.* Arlington, VA: Author.

Darling-Hammond, L. (2000). *Studies of excellence in teacher education.* Washington, DC: AACTE Publications.

Doyle, W. (1983). Academic work. *Review of Educational Research, 53,* 159–199.

Dwyer, C. A. (1993). Teaching and diversity: Meeting the challenge for innovative teacher assessments. *Journal of Teacher Education, 44,* 119–129.

Dwyer, C. A. (1994). *Development of the knowledge base for PRAXIS III: Classroom performance assessment criteria.* Princeton, NJ: Educational Testing Service.

Dwyer, C. A. (1998). Psychometrics of Praxis III: Classroom performance assessment. *Journal of Personnel Evaluation in Education, 12,* 163–187.

Englert, C. S., Tarrant, K. L., & Mariage, T. V. (1992). Defining and redefining instructional practice in special education: Perspectives on good teaching. *Teacher Education and Special Education, 15,* 62–86.

Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record, 107,* 186–213.

Fuchs, L. S., Fuchs, D., & Bishop, N. (1992). Teacher planning for students with learning disabilities: Differences between general and special educators. *Learning Disabilities Research & Practice, 7,* 120–128.

Goldhaber, D., & Anthony, E. (2004). *Can teacher quality be effectively assessed?* Seattle, WA: University of Washington, Center on Reinventing Public Education. Retrieved September 2, 2004, from http://www.crpe.org

Good, T. L. (1979). Teacher effectiveness in the elementary school. *Journal of Teacher Education, 30*(2), 52–64.

Greenwood, C. R., & Maheady, L. (1997). Measurable change in student performance: Forgotten standard in teacher preparation? *Teacher Education and Special Education, 20,* 265–275.

Haager, D., Gersten, R., Baker, S., & Graves, A. (2003). The English-language learner classroom observation instrument: Observations of beginning reading instruction in urban schools. In S. Vaughn & K. Briggs (Eds.), *Reading instruction observation tools* (pp. 111–144). Baltimore: Brookes.

Hess, F. M. (2001). *Tear down this wall: The case for a radical overhaul of teacher certification.* Washington, DC: Progressive Policy Institute. Retrieved May 11, 2004, from: http://www.ppionline.org/documents/teacher_certification.pdf

Humphrey, D. C., & Weschler, M. E. (2005, September). *Insights into alternative certification: Initial findings from a national study.* Retrieved February 7, 2006, from http://www.sri.com/policy/cep/pubs/teachers/AltCert_TCR_article.pdf

Interstate New Teacher Assessment and Support Consortium. (2001). *Model standards for licensing general and special education teachers of students with disabilities: A resource for state dialogue.* Washington, DC: Author.

Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research, 62,* 129–169.

Kennedy, M. M. (1996). Research genres in teacher education. In F. B. Murray (Ed.), *The teacher educator's handbook: Building a knowledge base for the preparation of teachers* (pp. 120–152). San Francisco: Jossey-Bass.

Kennedy, M. M. (1999). Approximations to indicators of student outcomes. *Educational Evaluation and Policy Analysis, 21,* 345–363.

Medley, D. M. (1978). Research in teacher effectiveness: Where it is and how it got here. *Journal of Classroom Interaction, 13*(2), 16–21.

Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance: An empirical approach.* New York: Longman.

National Commission on Excellence in Elementary Teacher Preparation for Reading Instruction. (2003). *Prepared to make a difference.* Newark, DE: International Reading Association.

Nevin, A., Thousand, J., Parsons, A. S., & Lilly, M. S. (2000). The kids keep me fresh! Results of a follow-up survey of graduates from two special education teacher education programs who are teaching. *Teacher Education Quarterly, 27,* 87–108.

No Child Left Behind Act of 2001, 20 U. S. C. 70 § 6301 *et seq.* (2002).

Nougaret, A. A., Scruggs, T. E., & Mastropieri, M. A. (2005). Does teacher education produce better special education teachers? *Exceptional Children, 71,* 217–229.

Nowacek, E. J., & Blanton, L. P. (1996). A pilot study investigating the influence of a collaborative methods course on preservice elementary teachers. *Teacher Education and Special Education, 19,* 298–312.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62,* 307–332.

Perie, M., Grigg, W., & Donahue, P. (2005). *The nation's report card: Reading 2005* (NCES 2006-451). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Reynolds, A. (1992). What is competent beginning teaching? A review of the literature. *Review of Educational Research, 62,* 1–35.

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes.* Washington, DC: Economic Policy Institute.

Rosenberg, M. S., Boyer, K. L., Sindelar, P. T., & Misra, S. (in press). Alternative route programs to certification in special education: What we know about program design, instructional delivery, and participant characteristics. *Exceptional Children.*

Rosenberg, M. S., & Sindelar, P. T. (2005). The proliferation of alternative routes to certification in special education: A critical review of the literature. *The Journal of Special Education, 39,* 117–127.

Rosenshine, B. V. (1986). Synthesis of research on explicit teaching. *Educational Leadership, 43*(7), 60–69.

Schools and Staffing Survey. (n.d.). *SASS teacher questionnaire.* Retrieved March 13, 2003, from http://www.nces.ed.gov/surveys/SASS/question9900.ASP#Teacher

Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3–36). New York: Macmillan.

Sindelar, P. T., Daunic, A., & Rennells, M. S. (2004). Comparisons of traditionally and alternatively trained teachers. *Exceptionality, 12,* 209–223.

Sindelar, P. T., Smith, M. A., Harriman, N. E., Hale, R. L., & Wilson, R. J. (1986). Teacher effectiveness in special education programs. *The Journal of Special Education, 20,* 195–207.

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher, 9*(11), 11–16.

Stanovich, P. J., & Jordan, A. (1998). Canadian teachers' and principals' beliefs about inclusive education as predictors of effective teaching in heterogeneous classrooms. *Elementary School Journal, 98,* 221–238.

Study of Personnel Needs in Special Education. (n.d.). *SPeNSE instruments.* Retrieved March 1, 2003, from http://ferdig.coe.ufl.edu/spense/instruments.html

Teacher Follow-up Survey. (n.d.). *Current teacher questionnaire.* Retrieved February 5, 2006, from http://nces.ed.gov/surveys/SASS/question05.asp

Tulbert, B., Sindelar, P. T., Correa, V. I., & La Porte, M. M. (1996). Looking in the rear view mirror: A content analysis of teacher education and special education. *Teacher Education and Special Education, 19,* 248–261.

Vandevoort, L. G., Amrein-Beardsley, A., & Berliner, D. C. (2004). National board certified teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46). Retrieved September 8, 2004, from http://epaa.asu.edu/epaa/v12n46/

Walsh, K. (2001). *Teacher education reconsidered: Stumbling for quality.* Retrieved January 15, 2001, from the Abell Foundation Web site: http://www.abell.org/pubsf.htm

Wideen, M., Mayer-Smith, J., & Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research, 68,* 130–178.

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education, 53,* 190–204.